International Journal of Advanced Research in Engineering and Technology (IJARET)
Volume 12, Issue 11, November 2021, pp.48-62 Article ID: IJARET_12_11_006
Available online at https://iaeme.com/Home/issue/IJARET?Volume=12&Issue=11
ISSN Print: 0976-6480 and ISSN Online: 0976-6499
DOI: https://doi.org/10.34218/IJARET.12.11.2021.006

ANALYZING AND IDENTIFYING FAKE NEWS USING ARTIFICIAL INTELLIGENCE

Parth Arora[1], Aniket Anvekar[2], Heetanshu Raval[3], Dr. Rohit B Patel[4]

[1]Maharaja Surajmal Institute of Technology, Guru Gobind Singh Indraprastha University, Delhi, India

[2]Rizvi College of Engineering, Mumbai University, Mumbai, India

[3]U.V Patel College of Engineering, Ganpat University, Mehsana, India

[4]Adani Institute of Infrastructure Engineering, Ahmedabad, India

ABSTRACT

The main reason behind the spread of fake news is because of many fake and hyperpartisan sites present on the Internet. These fake sites try to manipulate the truth which creates misunderstanding in society. Therefore, it is important to detect fake news and try to make people aware of the truth. This paper gives an insight into how to detect fake news using Machine Learning and Deep Learning Techniques. On observing our data, we have categorized our data into five attributes namely Title, Text, Subject, Date, and Labels. In order to develop an efficient fake news detection system, the feature along with its degree of impact on the system must be taken into consideration. This paper attempts at providing a detailed analysis of detecting fake news using various models such as LSTM, ANN, Naïve Bayes, SVM, Logistic Regression, XGBoost, and Bert.

Key words: Hyperpartisan, Machine Learning, Deep Learning, NLP, TfidfVectorizer, Naïve Bayes, Support Vector Machine, LSTM, ANN, XG Boost, Bert.

Cite this Article: Parth Arora, Aniket Anvekar, Heetanshu Raval and Rohit B Patel, Analyzing and Identifying Fake news using Artificial Intelligence, *International Journal of Advanced Research in Engineering and Technology (IJARET)*, 12(11), 2021, pp. 48-62.
https://iaeme.com/Home/issue/IJARET?Volume=12&Issue=11

1. INTRODUCTION

There has been an upsurge in the number of bogus websites on the Internet in recent years. Fake news and a lack of faith in the media are creating problems in our society that have far-reaching consequences. While the volume of fake news on the internet appears to be increasing, some people may find it progressively troublesome to tell what is true and what is not. As the 2016 Presidential election demonstrated, a lack of information literacy can have real-world consequences. It is important to establish skills to recognize and critically examine news you read on social media and other platforms because it can be quite compelling [1].

Fake news campaigns are a type of modern information warfare that nations and other entities undertake to weaken their opponents' authority and legitimacy. Individuals who check the accuracy of published news are known as fact-checkers. Those experts deconstruct fake news by pointing out its inaccuracies. Traditional fact-checking can be supplemented by machine learning and Natural Language Processing (NLP), according to research.

Fake news identification is a difficult undertaking due to the size of the system and the intricacies of the forces that influence it. The parameters that influence the development of an effective false news detecting system are critical. When reading news on the Internet, people must evaluate a variety of elements such as the news article's source, the date and time, the author's name, and a variety of other aspects that influence fake news. These aspects that influence fake news are dynamic in nature, and the selection of prominent traits is crucial in forecasting an accurate result [2].

Artificial intelligence is a conventional term that alludes to procedures that empower PCs to impersonate human conduct. All of this is made conceivable through AI, which is a bunch of calculations that have been educated on data.

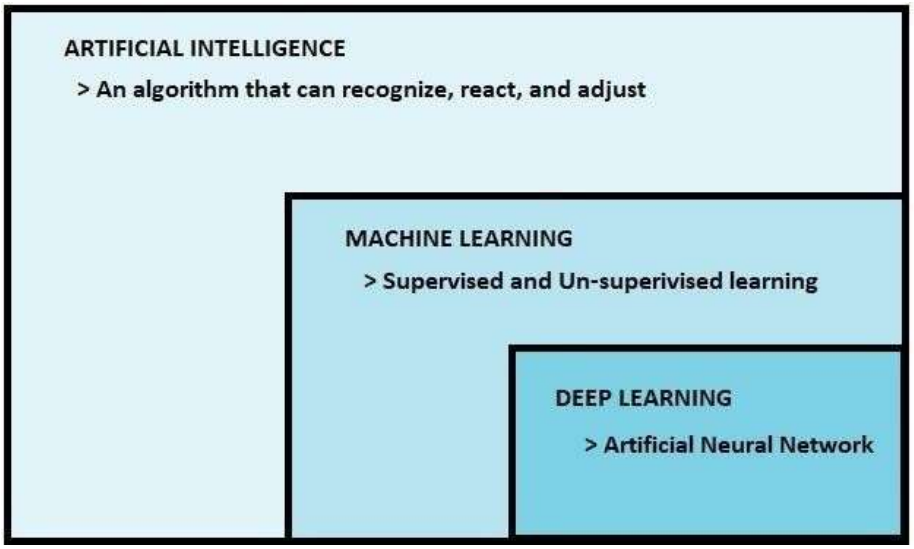


Figure 1

Deep Learning, on the other hand, is a sort of Machine Learning that is inspired by the human brain's structure. Deep learning algorithms analyze data with a predetermined logical structure in order to reach similar conclusions as humans. Deep learning does this by employing a multi-layered structure of algorithms known as "Neural Networks".

The neural network's plan is roused by the construction of the human cerebrum. Neural networks can be educated to play out the very errands on information that our minds do when distinguishing designs and arranging various kinds of data.

The autonomous neural networks layers can likewise be considered as a sort of channel that works from extensive to unobtrusive and expands the likelihood of perceiving and producing a right outcome. The brain of human beings works likewise. At whatever point we get new data, the mind attempts to contrast it and known elements. Deep neural networks make use of the same notion.

Various assignments like clustering, classification, or regression can be accomplished with the assistance of an Artificial Neural Network. These networks permit us to gather or sort untagged information dependent on likenesses between tests in that information. What's more, on account of order, the networks can be trained on marked datasets to characterize the examples in that dataset into various classifications which can be either paired or multiclass.

As a rule, neural networks can execute similar assignments as customary ML methods. But, vice versa is not true.

Deep learning is responsible for all recent developments in artificial intelligence. Auto-driven cars, IM Bot, and virtual assistants like Google assistant and Siri would not exist without deep learning. Amazon Prime or Disney+Hotstar would have no notion about what movies or TV shows we like to watch or not, and the Google Translate app would remain as basic as it was a decade ago (before Google transitioned to neural networks for this app). Neural networks are at the heart of all of these innovations.

We could even go so far as to suggest that deep learning and ANN are driving a new industrial revolution today.

Deep learning is, by the day's end, awesome and most clear way to deal with genuine machine knowledge we have so far.

Conventional machine learning methods were mainly used, before deep learning came into action. Examples of Machine Learning methods such as Decision Trees, SVM, Naïve Bayes Classifier, Logistic Regression and XG Boost, will further be discussed and compared with deep learning models like LSTM, ANN and Bert [3].

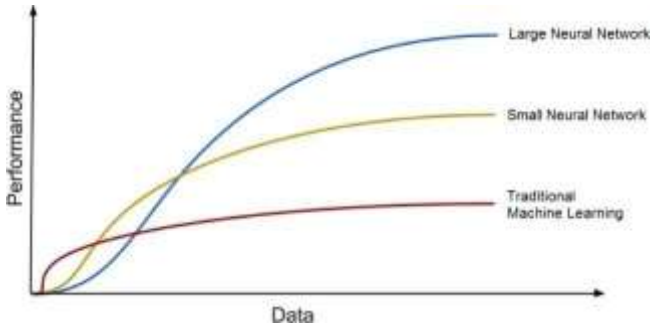


Figure 2

Conventional machine learning techniques such as SVM and Naïve Bayes classifiers limit their improvisation beyond a saturation point, however Deep Learning models improve with increasing amounts of training data [4].

2. BACKGROUND AND RELATED WORK

Prior to giving the subtleties of the proposed Deep Learning and Machine Learning Algorithms, we present the problem statement and related works [5].

2.1 Problem Statement

The use of online newspaper publishing platform for news consumption has two sides. From one perspective, consumers seek out and consume news via this platform because of its low cost, easy accessibility, and fast transmission of data. Then again, it facilitates the endless dispersion of “fake news" or inferior quality news that contains intentionally deceptive material. The pervasive spreading of fake news has the potential to have tremendously detrimental results for both individuals and society. As a result, detecting fake news has recently become an emergent study topic that is becoming a center of attention.

First, fake news is purposefully written which allows readers to believe an information which is not true, which in turn makes it difficult to detect fake news on the basis of news content; as a result, we must include auxiliary information, such as user social engagements on the online news platform, to assist in making a decision. Second, utilizing this auxiliary data is

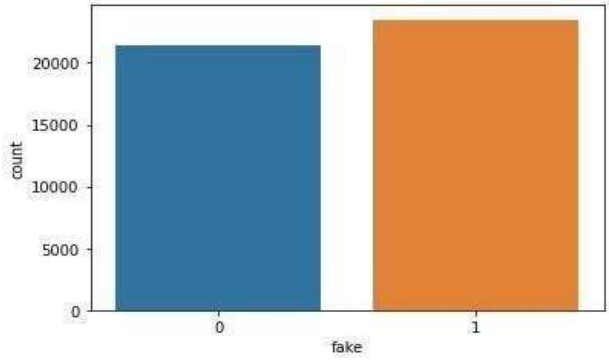
difficult in and of itself, as consumers' social interactions with fake news generate large, incomplete, unstructured, and noisy data.

2.2 Dataset

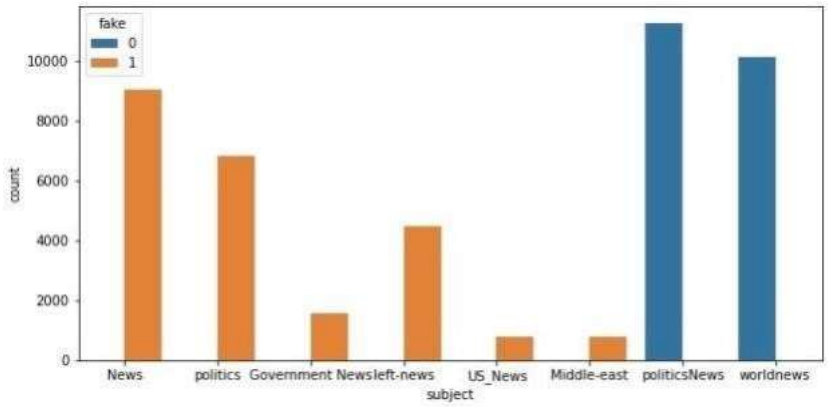
The factors identified which influence the fake news share a varying deal of change .In order to tackle with fake news we have used a dataset available on Kaggle which consist of five attributes namely Title, Text, Subject, Date and Label .Title contains the news heading whereas the Text contains the details about the news. The label conveys whether the news is true or fake. Subject contains the category in which the news belongs to, for example - political, leftist or rightist. Date contains the Date of article published. A fake news detection system is designed considering these factors.

	title	text	subject	date	fake
0	As U.S. budget fight looms, Republicans flip t...	WASHINGTON (Reuters) - The head of a conservat...	politicsNews	December 31, 2017	0
1	U.S. military to accept transgender recruits o...	WASHINGTON (Reuters) - Transgender people will...	politicsNews	December 29, 2017	0
2	Senior U.S. Republican senator: 'Let Mr. Mueli...	WASHINGTON (Reuters) - The special counsel inv...	politicsNews	December 31, 2017	0
3	FBI Russia probe helped by Australian diplomat...	WASHINGTON (Reuters) - Trump campaign adviser ...	politicsNews	December 30, 2017	0
4	Trump wants Postal Service to charge 'much mor...	SEATTLE/WASHINGTON (Reuters) - President Donal...	politicsNews	December 29, 2017	0
...	...	...	...	...	...
21412	'Fully committed' NATO backs new U.S. approach...	BRUSSELS (Reuters) - NATO allies on Tuesday we...	worldnews	August 22, 2017	0
21413	LexisNexis withdrew two products from Chinese ...	LONDON (Reuters) - LexisNexis, a provider of l...	worldnews	August 22, 2017	0
21414	Minsk cultural hub becomes haven from authorities	MINSK (Reuters) - In the shadow of disused Sov...	worldnews	August 22, 2017	0

Figure 3 Snapshot of our Dataset



(a)



(b)

Figure 4 Distribution of data

2.3 Methodology Used

2.3.1 Natural Language Processing (NLP)

Text mining is utilized to recognize realities, connections, and cases that would somehow stay hid in a huge volume of literary large information. When extricated, this data is converted into an organized structure that can be examined further to get significant experiences. Text mining utilizes a variety of procedures to deal with text, perhaps the most significant and generally utilized technique is Natural Language Processing.

Natural language Processing (NLP) is the capacity of program to decipher communicated in and composed human language, often known as natural language. It's an artificial intelligence component. Natural language processing (NLP) allows PCs to understand natural language in the same manner that humans do. Artificial intelligence is used in natural language processing to absorb real-time data, elucidate it, and make sense of it in a way that a computer can understand, regardless of the language being spoken or written. Computers have programs to read and microphones to gather audio, much as people have different sensors such as ears to hear and eyes to see. The input is changed to code that the computer can interpret during the processing [6].

There are two primary stages to NLP: preprocessing of data and algorithm development.

Data Preprocessing includes preparing and "cleaning" text information for machines to have the option to investigate it. preprocessing places information in functional structure and features highlights in the text that can be worked with an algorithm. There are multiple ways this should be possible, including:

- Tokenization - This is when text is separated into more modest units to work with. Stop word expulsion - when normal words are eliminated from text so special words that offer the most data about the text remain.
- Lemmatization - Its objective is to reduce the word to its base form and then grouping different forms of the word together.
- Stemming- This partitions words with expression in them to root structures.

An algorithm is created to deal with data when the data has been processed. There are a wide range of natural language processing algorithms, however two principle types are ordinarily utilized:

Machine Learning-Based System. ML algorithms involves defining phenomena in terms of numbers and then using the numbers to either imply or deduce cause and effect. They figure out how to manage responsibilities dependent on training data given to them, and as more data processes, they modify their methodologies. Natural language processing algorithms refine their own standards through continued learning and processing , utilizing a blend of ML, deep learning, and neural networks.

Deep Learning Based System

```
input_sent
' Donald Trump Sends Out Embarrassing New Year's Eve Message; This is DisturbingDonald Trump just coul

output_sent
'donald trump send embarrass new year eve messag disturbingdonald trump wish american happi new year '
```

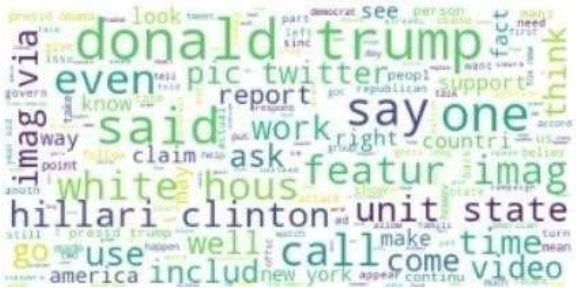


Figure 5

2.3.2 Naïve Bayes Classifier

It is a classification method based on Bayes' theorem and the assumption of independent predictor variables. In simple terms, the Naïve Bayes classifier assumes that the existence of a feature in a class has nothing to do with the existence of another feature.

For example, if a vegetable is wine-red, round and about 3 inches in diameter, it is called a potato. Even if these traits are interdependent or dependent on the existence of other traits, they all increase the possibility that this fruit is an apple, which is why it is called "Naïve".

The Naïve Bayes model is easy to carry out and is uncommonly successful for huge datasets. Naïve Bayes is prestigious to beat even the most developed order frameworks because of its effortlessness [7]. The Bayes theorem helps in calculating posterior probability (the revised or updated probability of an event occurring after taking into consideration new information) P(c|x) from P(c), P(x), and P(x|c) with P(c), P(x), and P(x|c). Check out the following equation:

$P(c|x) = [P(x|c) . P(c)] / P(x)$

Where,

P(c|x)= Posterior Probability

P(x|c)= Likelihood

P(c)= Class Prior Probability P(x) = Predictor Prior Probability

2.3.2.1 Gaussian Naïve Bayes

Gaussian Distribution conveys ceaseless qualities related with each component. A Gaussian distribution is likewise called Normal distribution.

2.3.2.2 Multinomial Naïve Bayes

The recurrence with which explicit events were made by a multinomial distribution is addressed by feature vectors. This is the most well-known event model for record grouping.

2.3.3 Logistic Regression

Logistic Regression comes under the Supervised Learning approach which is the most widely used Machine Learning algorithm. It is a model for predicting a categorical dependent variable out of a number of independent variables.

As a result, the output must be a discrete or categorical value. It can be any binary value like Yes or No, 0 or 1, true or false, and so on. Rather than giving exact values like 0 and 1, it gives probabilistic values which are and far between 0 and 1[8].

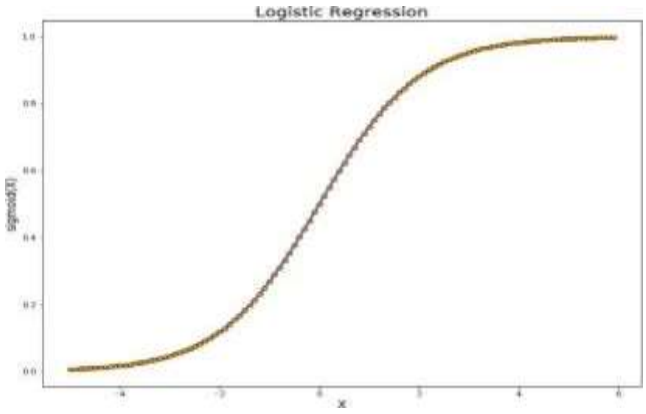


Figure 6

Apart from how they are implemented, Logistic Regression is really similar to Linear Regression. For regression problems, Linear Regression is employed, while for classification difficulties, Logistic Regression is used.

2.3.4 Support Vector Machine

The "Support Vector Machine" is a type of supervised Machine Learning algorithm which is used in dealing with classification and regression problems. In most cases, it is used to classify output labels. Each input data value is a point in an n-dimensional space vector(where n is the quantity of distinct features we have inside the dataset), with the value of each feature equaling the value of a single coordinate. After that, we perform classification by identifying the hyperplane that effectively separates the two groups [9].

Simply put, support vector is the coordinate of each observation. The SVM classifier is the most effective boundary separating the two classes (hyper-plane/line).

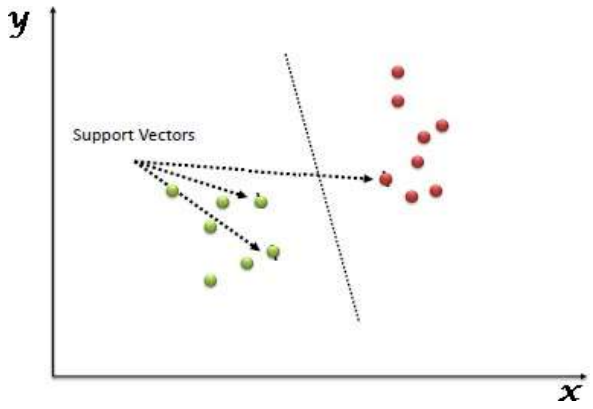


Figure 7

2.3.5 XGBoost

XGBoost is a decision tree-based troupe Machine Learning algorithm that uses a gradient boosting framework. In estimate issues including unstructured data (text, pictures etc.) neural networks will beat some other AI based calculation. By the by, when we are working with little to-medium organized or even information, tree-based algorithms are productive and vigorous [10].

2.3.6 Long Short-Term Memory (LSTM)

RNNs are really a kind of neural network that permits past step outputs to be employed as inputs to the current step. RNN is very well fitted for sequential prediction applications because of this characteristic. However, due to the vanishing gradient problem which is caused by gradient propagation in the recurrent network, these networks show weakness in learning long-term dependency on massive inputs. Long Short-Term Memory networks or "LSTM" are a form of RNN that could learn semi-permanent dependencies. They perform incredibly well on a broad variety of conditions and are currently in widespread use. LSTMs are specifically developed to prevent the problem of semi- permanent reliance. It is not one of their struggles to be told that they memorize information for long periods of time. All repeated neural networks take the form of a sequence of neural network continuation modules [11].

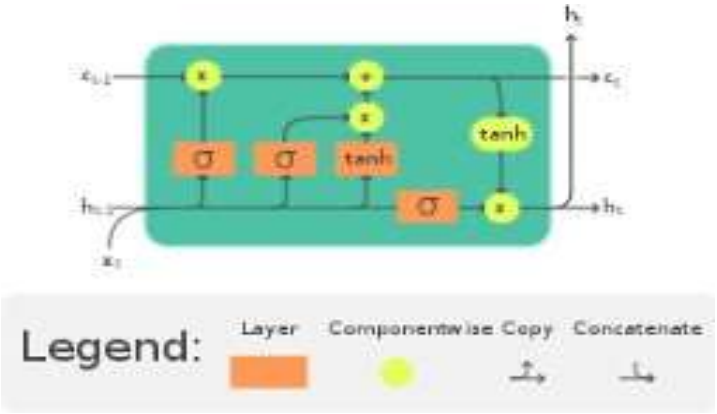


Figure 8

Each conventional neuron in the concealed layer of a regular RNN is replaced by an LSTM, which is a memory unit. Figure 8 shows the structure of an LSTM memory cell. The flow of information to and fro from the cell is controlled by an input gate, a forget gate, and an output gate in the LSTM unit.

2.3.7 Artificial Neural Network (ANN)

An ANN is computing model which could be used to complete tasks like prediction, classification, and decision- making. It is made up of synthetic neurons. These synthetic neurons are exact replicas of human brain neurons. The signals to conduct the activities are sent by neurons in the brain. Artificial neurons link in a neural network to complete tasks in a similar way. Weight refers to the strength of the connection between the artificial neurons [12].

- Since several connections between input and output are non- linear, ANN can understand and model non-linear and complex connections.
- After training, ANN may infer previously unknown relationships from previously unknown data, making it generic.
- Unlike many other machine learning models, ANN has no requirements for datasets, such as Gaussian distribution or any other distribution.

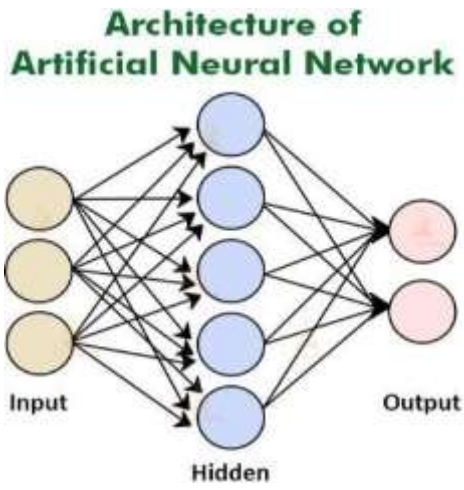


Figure 9

2.3.8 Bert

BERT, or Bidirectional Encoder Representations from Transformers, enhances normal Transformers by eliminating the unidirectionality prerequisite by pre-preparing with a masked language model (MLM). This language model veils a few tokens from the contribution at irregular, fully intent on anticipating the covered word's unique jargon is dependent on its specific situation. The MLM point, dissimilar to left-to-right language model pre-training, permits the portrayal to coordinate the left and right settings, permitting us to pre-train a deep bidirectional Transformer. BERT likewise utilizes a next sentence forecast task, which pre-trains text-pair portrayals couple with the masked language model [13].

Pre-training and tuning are the two cycles in BERT. The model is prepared on unlabelled information across numerous pre-training undertakings during pre-training. The BERT model is calibrated by first introducing it with the pre-trained boundaries and afterward adjusting every one of the boundaries utilizing marked information from the downstream positions. Despite the fact that the models are beginning with similar pre-trained boundaries, each downstream assignment has its own calibrated model.

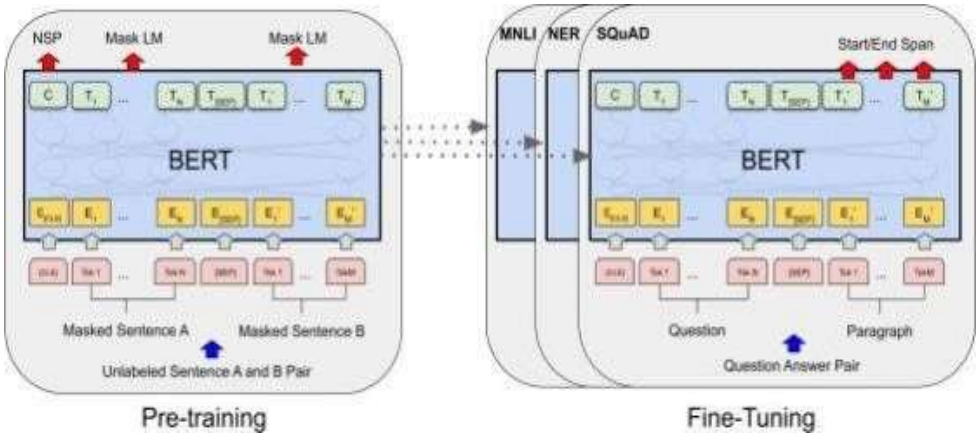


Figure 10

3. PROPOSED MODELS AND RESULTS

In this section, we will look at the results, that is, F1 score, Accuracy, Recall score and Area under the ROC curve, that is, ROC AUC. Also, the Confusion Matrix for every model are presented in their respective columns.

TF-IDF: “Term Frequency - Inverse Document Frequency”. This is a procedure to measure a word in archives, we for the most part figure a load to each word which means the significance of the word in the report and corpus.

This technique is an extensively used strategy in Data Retrieval and Text Mining.

TF - The measure of times a term shows up in a textual report.

IDF - A proportion of whether a term is uncommon or normal in the text

TF-IDF = Multiplication (TF[n, d], IDF[n]) where, n – count of corpus d - document

3.1 Naïve Bayes Classifier

3.1.1 Gaussian Naïve Bayes

```
**********Naive Bayes (TfidfVectorizer)**********
F1 score: 0.9795106769257371
Accuracy: 0.96208
ROC AUC: 0.9526970974768894
```

Figure 11

Figure 11 shows the results of Gaussian Naïve Bayes Classifier.

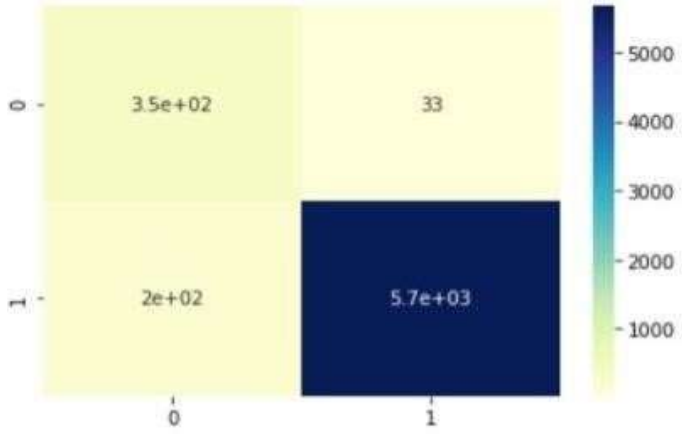


Figure 12 Confusion Matrix for Gaussian Naïve Bayes

3.1.2 Multinomial Naïve Bayes

```
**********Multinomial Naive Bayes (TfidfVectorizer)**********
F1 score: 0.9890982844587172
Accuracy: 0.97936
ROC AUC: 0.9526970974768894
```

Figure 13

Figure 13 shows the results of Multinomial Naïve Bayes Classifier

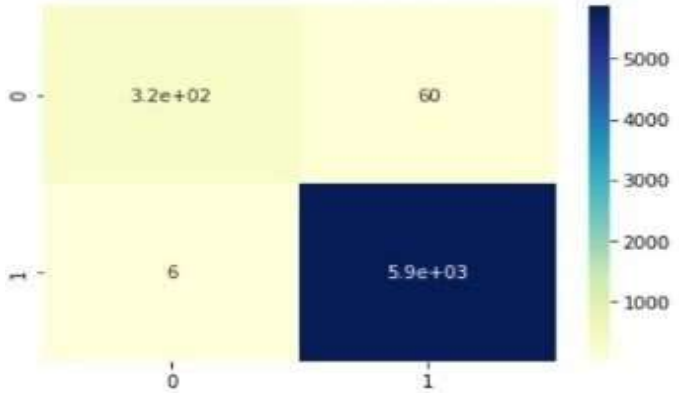


Figure 14 Confusion Matrix of Multinomial Naïve Bayes

3.2 Logistic Regression

```
**********Logistic Regression (TfidfVectorizer)**********
F1 score: 0.9944029850746268
Accuracy: 0.98944
ROC AUC: 0.9526970974768894
```

Figure 15

Figure 15 shows the results of Logistic Regression

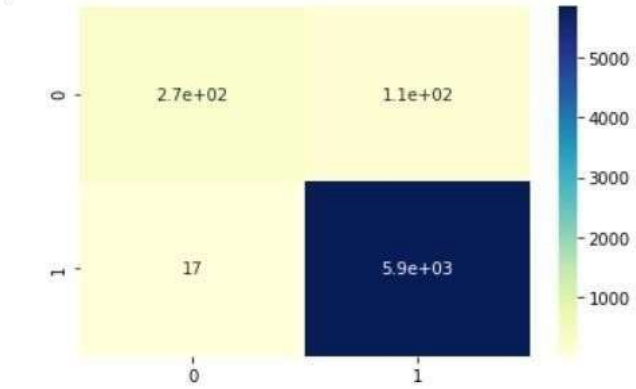


Figure 16 Confusion Matrix for Logistic Regression.

3.3 Support Vector Machine (SVM)

```
**********SVM (TfidfVectorizer)*********
F1 score: 0.9875321249255382
Accuracy: 0.98456
ROC AUC: 0.9898000974022053
```

Figure 17

Figure 17 shows the results of Support Vector Machine

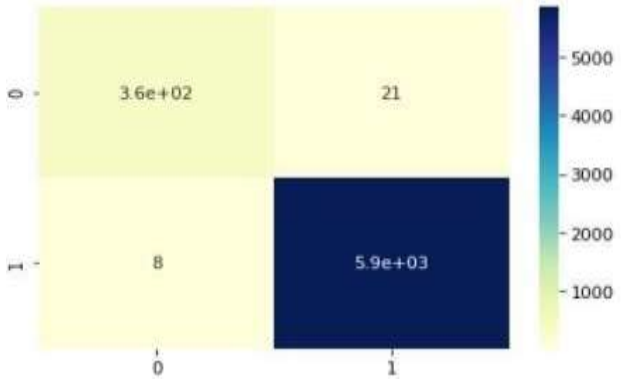


Figure 18 Confusion Matrix for SVM

3.4 XGBoost

```
********** XGBoost (TfidfVectorizer)**********
F1 score: 0.9888066825775657
Accuracy: 0.98776
ROC AUC: 0.9891163142433062
```

Figure 19

Figure 19 shows the results of XGBoost

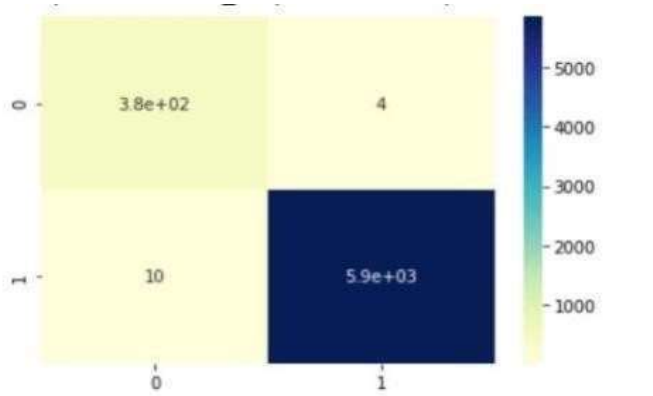


Figure 20 Confusion Matrix for XGBoost

3.5 Long Short Term Memory (LSTM)

```
[32] model.summary()

Model: "model_LSTM"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
input_1 (InputLayer)         [(None, 1)]               0
_________________________________________________________________
text_vectorization (TextVect (None, 418)               0
_________________________________________________________________
embedding (Embedding)        (None, 418, 128)          1280000
_________________________________________________________________
lstm (LSTM)                  (None, 64)                49408
_________________________________________________________________
dense (Dense)                (None, 1)                 65
=================================================================
Total params: 1,329,473
Trainable params: 1,329,473
Non-trainable params: 0
_________________________________________________________________
```

```
Accuracy Score of LSTM model: 0.9902004454342984
Recall Score of LSTM model: 0.9856068743286789
Precsion Score of LSTM model: 0.9954436971143416
f1 Score of LSTM model: 0.9905008635578585
```

Figure 21

Figure 21 shows the results of LSTM.

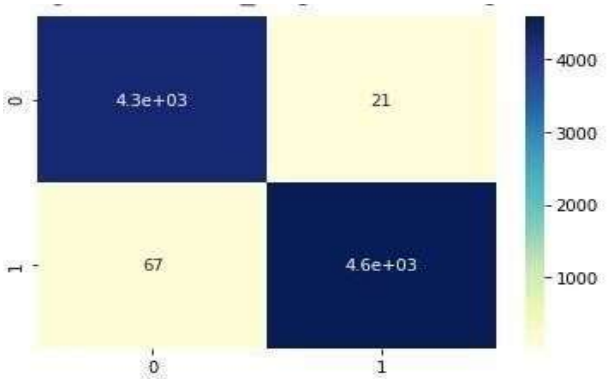


Figure 22 Confusion Matrix for LSTM

3.6 Artificial Neural Network (ANN)

Printing classification_report for Test Set

	precision	recall	f1-score	support
0	0.97	0.97	0.98	1406
1	0.97	0.97	0.98	2794
accuracy			0.98	4200
macro avg	0.98	0.98	0.98	4200
weighted avg	0.98	0.98	0.98	4200

Figure 23

Figure 23 shows the results of Artificial Neural Network

```
Epoch 1/3
49/49 [==============================] - 6s 110ms/step - loss: 0.1762 - accuracy: 0.9416 - val_loss: 0.0748 - val_accuracy: 0.9844
Epoch 2/3
49/49 [==============================] - 5s 108ms/step - loss: 0.0414 - accuracy: 0.9920 - val_loss: 0.0536 - val_accuracy: 0.9884
Epoch 3/3
49/49 [==============================] - 5s 107ms/step - loss: 0.0184 - accuracy: 0.9975 - val_loss: 0.0458 - val_accuracy: 0.9898
```

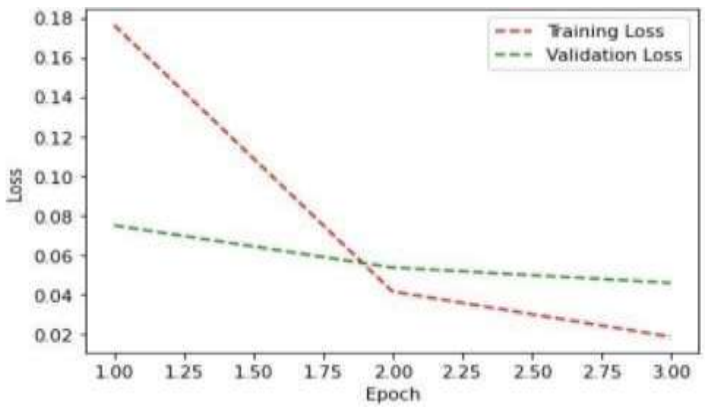


Figure 24 Training Loss vs Validation Loss

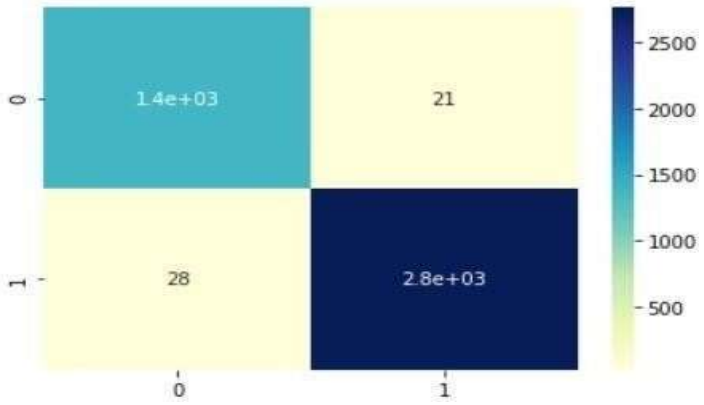


Figure 25 Confusion Matrix for ANN

3.7 Bert

```
    news_bert_report(pred_lis.detach().numpy(), label_lis.detach().numpy())

f1-score:  0.9950845222395396
accuracy:  0.9946980473296263
```

Figure 26

Figure 26 shows the results of Bert

4. RESULT AND OUTPUT

Table 1

Models Used	Accuracy
Gaussian Naïve Bayes	96.21%
Multinomial Naïve Bayes	97.94%
Logistic Regression	98.94%
SVM	98.45%
XGBoost	98.77%
ANN	98.00%
LSTM	99.02%
Bert	99.47%

5. CONCLUSION

Owing to the growth of emerging technologies, the rate of using online news platforms has increased rapidly. Nowadays, people are relying on online news publishing platforms to get quick insights on what is trending, and this has led to a decrease in the use of news channels and newspapers. Due to the lack of stringent policies over the Internet, there has been a significant increase in fraud and bogus news. The spread of fake news may lead to widespread destruction. Most of the Internet users are sharing news articles with their friends and families without checking the credibility of the news and this has resulted in unnecessary disputes between different societies and religious communities. In the future, we can't distinguish between real news and fake news as the rate of fake news has become disguise. In this digital

age, where hoax news is present everywhere on digital platforms, there is an ultimate need for fake news detection and this model serves its purpose by being the need of the hour tool. Fake News regarding sensitive topics leads to a toxic environment on the web. Fake News Detection is the systematic analysis of socially relevant data to distinguish whether it is real or fake

[15]. In this paper, we have compared various Machine Learning methods like Naïve Bayes, Logistic Regression, Support Vector Machine, XGBoost and Deep Learning Methods like ANN, LSTM and Bert. These methods were tested on a Kaggle dataset. On comparing the accuracies of all the above models, we can conclude that LSTM and BERT have out-performed all the Machine Learning models. In order to train a large volume of complex data, we propose a hybrid model which will help to attain higher accuracy.

Funding Statement: "The author(s) received no specific funding for this study."

Conflicts of Interest: "The authors declare that they have no conflicts of interest to report regarding the present study."

REFERENCES

[1] https://researchguides.uoregon.edu/fakenews "Fake News and Information Literacy"

[2] Akshay Jain and AmeyKasbe. "Fake News Detection." 2018 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS). Bhopal, India: IEEE. 2018.

[3] N. J. Conroy, V. L. Rubin and Y. Chen, Automatic deception detection: Methods for finding fake news, vol. 52, pp. 1-4, 2015.

[4] Uma Sharma, Sidarth Saran, Shankar M. Patil (Fake News Detection using Machine Learning Algorithms)

[5] Abdullah-All-Tanvir, Mahir, E. M., Akhter S., & Huq, M. R. (2019). Detecting Fake News using Machine Learning and Deep Learning Algorithms. 7th International Conference on Smart Computing & Communications (ICSCC), Sarawak, Malaysia, Malaysia, 2019, pp.1-5,

[6] Amirsina Torfi, Member, IEEE, Rouzbeh A. Shirvani, Yaser Keneshloo, Nader Tavaf, and Edward A. Fox, Fellow, IEEE (Natural Language Processing Advancements by Deep Learning: A Survey)

[7] Mykhailo Granik, Volodymyr Mesyura "Fake news detection using Naïve Bayes classifier" (IEEE)

[8] https://towardsdatascience.com/logistic-regression-detailed- overview-46c4da4303bc

[9] https://www.javatpoint.com/machine-learning-support-vector- machine-algorithm

[10] https://www.analyticsvidhya.com/blog/2018/09/an-end-to-end- guide-to-understand-the-math-behind- xgboost/

[11] Alex Sherstinsky (Fundamentals of Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) Network)

[12] https://techvidvan.com/tutorials/artificial-neural-network/

[13] Jacob Devlin, Ming-Wei Chang, Kenton Lee, Kristina Toutanova "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding". Available: https://arxiv.org/abs/1810.04805v2

[14] https://medium.com/@cmukesh8688/tf-idf-vectorizer-scikit- learn-dbc0244a911a

[15] Brewer P.R., Young D.G., Morreale M. "The impact of real news about fake news": Intertextual processes and political satire. International Journal of Public Opinion Research, 25 (2013), pp. 3